Naval Air Defense Threat Assessment: Cognitive Factors and Model

Michael J. Liebhaber, Ph.D.*

Pacific Science & Engineering Group, Inc.

6310 Greenwich Drive, Suite 200

San Diego, CA 92122

C.A.P. Smith, Ph.D.

Space and Naval Warfare Systems Center

Bldg. A33, Room 1403

San Diego, CA 92152-5001

Abstract

This paper reports the results of an investigation into the cognitive aspects of threat assessment. Threat assessment is the process of evaluating aircraft that are flying in the vicinity of one's ship and determining how much of a threat they represent to one's own ship and to the larger battle group. Data were collected from experienced U.S. Navy air defense officers as they interacted with a realistic scenario. Participants assigned threat and priority levels to selected aircraft, and described the factors and rationale that they used to arrive at their decisions. Twenty-two factors (e.g., Altitude, Speed, IFF) were identified. Participants consistently used different, but overlapping, subsets of factors, called *profiles*: A schema that lists the factors to be considered for a particular type of track, and the expected range of values that any given factor can take on. The participants appeared to sequentially evaluate the factors, and set threat levels that were inversely related to the fit between data and expectations. Aircraft that matched expectations were assigned lower threat levels than aircraft that did not match expectations. Threat Levels were biased by the Geopolitical Situation. A model was proposed that embodied the major findings in the data.

---

* This research was supported by the Office of Naval Research. Special thanks to Dr. Sue Hutchins at the Naval Postgraduate School in Monterey, CA for her assistance.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **2000**	2. REPORT TYPE	3. DATES COVERED **00-00-2000 to 00-00-2000**

4. TITLE AND SUBTITLE **Naval Air Defense Threat Assessment: Cognitive Factors and Model**	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Space and Naval Warfare Systems Center, San Diego,53560 Hull Street,San Diego,CA,92152-5001**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR’S ACRONYM(S) 11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**29**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

1. Introduction

This paper reports the results of a preliminary investigation into the process of aircraft threat assessment as it is conducted by experienced U.S. Navy Air Defense/Warfare officers. This study is part of an ongoing program of research into understanding and facilitating the assessment process. The goals of this study are to identify and describe the factors and cognitive processes that an Air Defense team onboard ship uses to assess and prioritize aircraft contacts, and to provide preliminary specifications for a cognitively-based threat assessment model. To support these goals, data were collected from experienced Naval AD personnel as they interacted with a realistic, computer-based scenario. Participants assigned threat and priority levels to selected aircraft in the vicinity of their ship, and described the factors and rationale that they used to arrive at their decisions.

Factors are the elements of data and information that are used to assess air contacts. Traditionally, they are derived from kinematics, tactical, and other data. Examples of such data include course, speed, IFF mode, and type of radar emitter. Threat assessment is defined, for use in this study, as the process of evaluating aircraft that are flying in the vicinity of one's ship, and determining how much of a threat they represent to the ship as well as to the battle group. It is one component in building and maintaining situation awareness. It includes gathering and reviewing relevant information, deciding what action should be taken, and how urgent that action needs to be.

2. Background

Aircraft contacts, called tracks, are processed by an automated combat information system, and displayed for the Air Defense team. The team must evaluate the displayed information, and decide what actions to take, if any. It is generally assumed that a standard set of factors or variables are used by Air Defense (AD) personnel, and that the process of evaluation and identification is handled by planned, predeterminded criteria. However, this assumption is driven more by historical precedence and doctrine, than by studies of use and reliance on data by AD team members. Therefore, this project was guided by the questions: What information do AD

personnel use when assessing aircraft contacts? Are some factors relied on more than others? and Do they use them systematically?

Current combat systems, such as those on Aegis ships, employ algorithms that rely on kinematics (e.g., range, course, speed, & altitude) and tactical (e.g., type of radar & recent track history) data to analyze and identify (ID) aircraft contacts. The system assigns an ID in conjunction with a human operator who is relying on a small set of factors (typically determined before the ship begins its deployment). Other relevant information, such as geography, intelligence reports, and country of origin, must be purposefully integrated by AD personnel to arrive at an assessment of threat level and priority. This integration requires a high level of tactical expertise that is not well documented or understood. In addition, there is a tacit assumption that the threat level of an aircraft is related to its identification (ID), such as Commercial or Tactical Aircraft, and therefore, the degree of threat posed by an aircraft can be methodically described by a predetermined set of factors.

Within the framework of Naturalistic Decision Making (NDM), the processes of threat assessment and situation awareness share many common elements. Situation awareness, as defined by Endsley [1995, p. 36], is the state of knowledge achieved by "… the perception of the elements in the environment, … the comprehension of their meaning, and the projection of their status in the near future." Most NDM models are characterized by a process that involves matching the pattern of a situation to sets of actions, and then selecting and evaluating an action with respect to goals and plans. For examples, see Endsley [1995], Klein [1993, 1997], Lipshitz and Shaul [1997], Marshall et al. [1996], Serfaty et al., MacMillan, Entin, and Entin [1997], Smith and Marshall [1997], and Zachary et al. [1992], and Zsambok and Klein [1997]. For models that emphasize metacognitive components, see Cohen et al. [1993], Cohen, Freeman, and Thompson [1995, 1997], Cohen and Klein [1995], and Federico [1997].

There is a great deal of overlap between the concepts in these models and the results of this project. It appears that threat assessment may be an instance of situation assessment, and that the process of threat assessment offers an insight into the process of building situation awareness.

3. Data Collection Methodology

3.1 Participants

Seven experienced Air Defense Officers participated in this study. They were selected from active duty Naval officers, who had the level of fleet and air defense experience that was necessary to understand and discuss the air threat assessment and prioritization process. Five had served as TAO at sea (avg. 14 months). Four had over 3 years of underway AD experience. Average total AD time was 6.5 years. Average underway AD time was 4.9 years.

3.2 Scenario

Several scenarios were developed for use in the Decision-making Evaluation Facility for Tactical Teams (DEFTT) laboratory at the Space and Naval Warfare (SPAWAR) Systems Center to study tactical decision making in demanding operational situations [Sonalysts, Inc., 1995]. The scenarios were developed using realistic air and surface tracks and associated operational events in a littoral, Persian Gulf situation. They were intended to place substantial mental demands on tactical decision makers while remaining representative of actual at-sea operations. Feedback from Navy officers with recent at-sea tactical experience indicated that the scenarios were demanding and realistic [Hutchins & Kowalski, 1993 as cited in Kelly & Moore, 1996].

Scenario “D” was selected for this study from the set of scenarios. It was chosen because it had a large proportion of critical tracks relative to total scenario run-time (15 minutes), was brief yet realistic, and engaged the interest of subjects in previous studies. The scenario depicted the USS Anzio (own ship) as it conducted a presence patrol approximately 50 miles off the coast of Kuwait. Six tracks were chosen for the participants to watch and evaluate. They represented a range of threat (very threatening to not threatening) and a variety of track types (jet, propeller, helicopter, military, & commercial), so as not to limit the range of comments by the participants to a particular type of track or situation.

The scenario interface is shown in Figure 1. The Geoplot is the map in the middle of the screen. Tracks are displayed on the Geoplot. The buttons on the left panel of the display allow participants to select tracks and airlane overlays, and to zoom in or out of the Geoplot. When a track is selected, its data (e.g., altitude, range, speed, etc.) appeared in the Character Read-out window on the right panel of the screen.

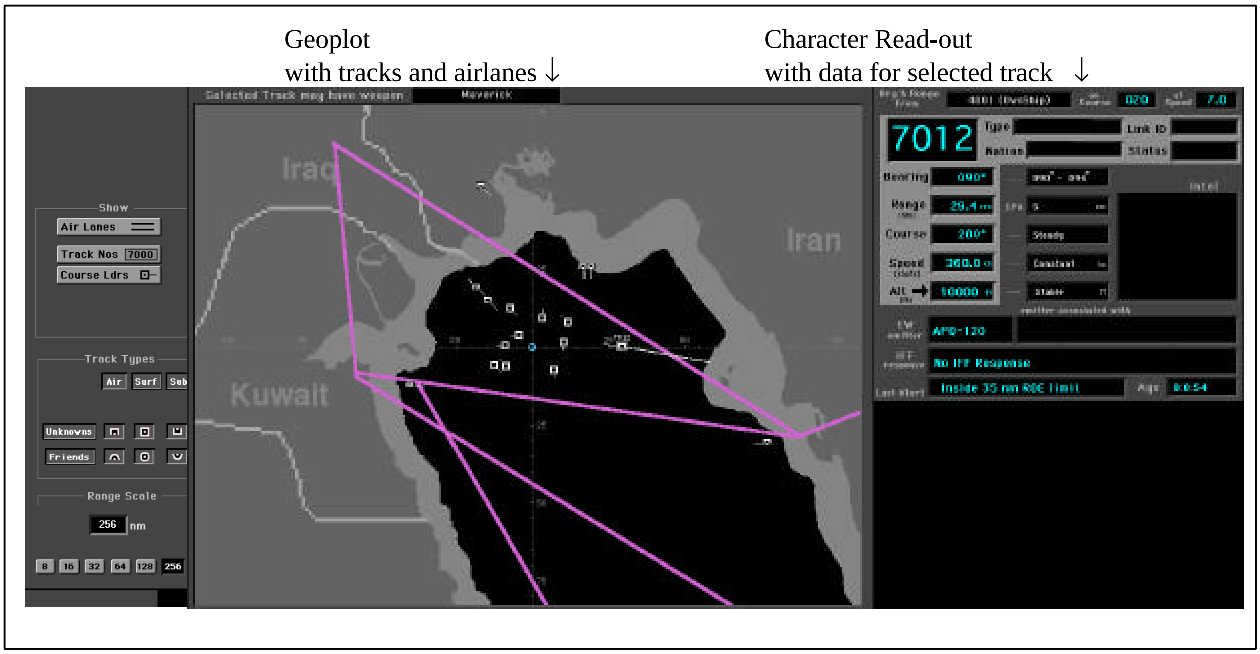

Figure 1. Scenario screen display.

3.3 Variables

Demographic Variables. Background information from the participants was collected, including experience (number of months of relevant, direct experience the participant has had in an Aegis CIC AD setting), rank, and time in rank.

Independent Variables. Participants read a Geopolitical Situation Briefing prior to viewing the scenario. It had two levels: Tense and Relaxed. The briefings were worded so that they gave the impression of either a Tense Situation (e.g., hostile actions have occurred and are expected to continue) or a Relaxed Situation (e.g., no recent hostilities, all factions relatively quiet). The briefings were adapted by a subject matter expert (SME) from the briefing for scenario “D”.

Briefing was a between-subjects variable, and participants were alternately assigned to the Tense or Relaxed version.

Outcome Measures. Data were collected at three equally spaced time intervals. Participants assigned Threat and Priority Level to six air tracks, and identified the factors that went into their decisions. The Levels were recorded by the computer, for each track at each time interval. Their spoken comments were recorded by audiotape for the entire time that they viewed the scenario; and later transcribed. Threat Level was defined as "How threatening, or capable of doing harm, is the track?" Priority was defined as "How much attention do you want to devote to this track?" The Levels ranged from 1 (Low) to 7 (High).

3.4 Procedure

An overview of the sequence of events is for each participant is:

- Participants read a Geopolitical Situation Briefing. They could refer to the briefings, as often as necessary, while they viewed the scenario. The experimenter started the experimental scenario when the participants were finished reading the briefing.
- The scenario paused automatically at scenario times 5, 10, and 15 minutes for data collection. The same six tracks were evaluated at each time interval. The rating scales appeared in the Data Entry Window on the far right of the screen. An example is shown in Figure 2. The Geoplot remained in view during the data collection process. Participants assigned a Threat and Priority Level for the Track by clicking on the Rating Scales. Lower on the scale meant less threat or priority, and higher meant more threat or priority. Participants could change either response. Responses were recorded only when participants clicked on the *Done* button.
- Participants described the Factors that they considered while making their decision.

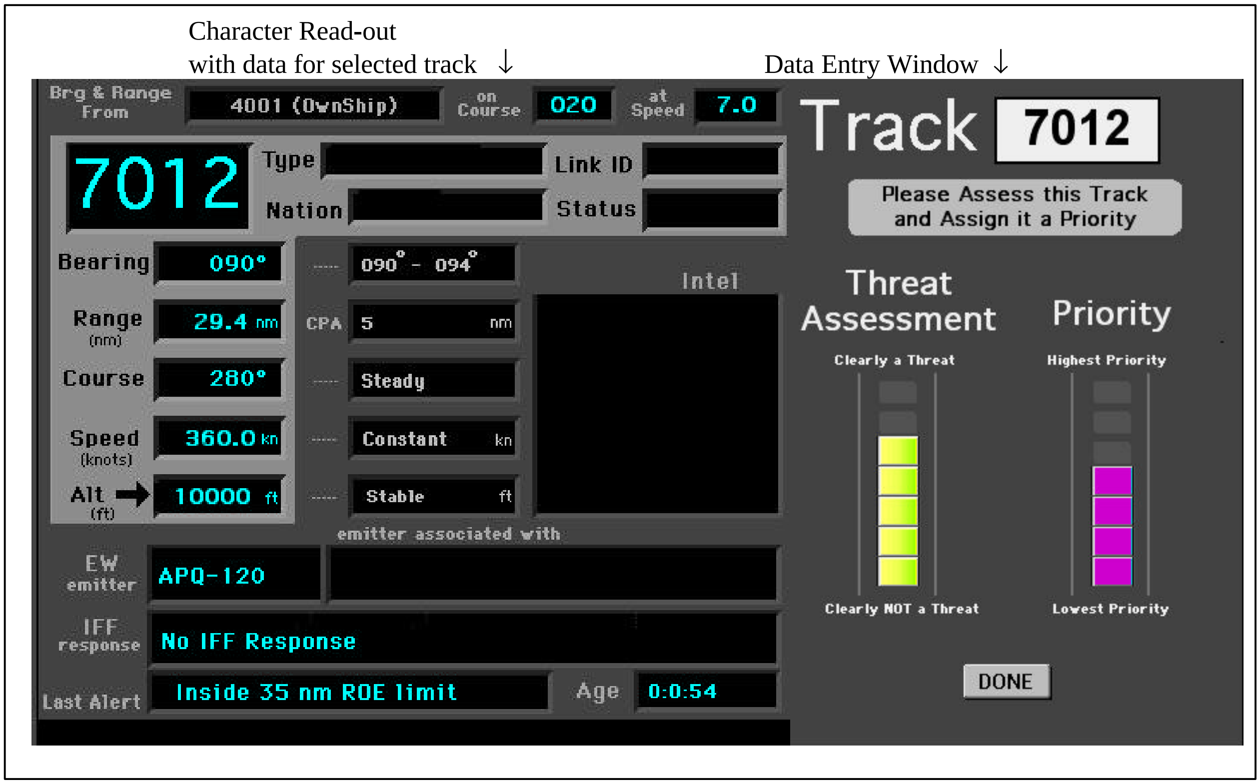

Figure 2. Scenario screen display with rating scales.

4.1 Preliminary Analysis: Tracks, Geopolitical Situation, and Experience

This analysis was conducted to answer questions pertaining to the scenario and Geopolitical bias, such as: Were the tracks sufficiently different from each other, and were participant's assessments relatively stable over the course of the scenario? and Did the Geopolitical Situation Briefing bias participant's responding? Two 2 (Briefing) X 6 (Track) X 3 (Time) ANOVAs were conducted. Tracks and Time (5, 10, & 15 minutes) were repeated within subjects. The dependent variables were Threat and Priority Level.

4.1.1 Scenario Tracks and Geopolitical Situation

Participants appeared to be sensitive to the unique factors in each track that they evaluated. They assigned different Threat Levels to different Tracks across all data collection intervals: Track main effect ($F(5,25) = 32.51$, $p < .01$), Track x Time ($F(10,50) = 12.40$, $p < .01$), and Briefing x Track x Time ($F(10,50) = 2.33$, $p = .024$) . There was also a high positive correlation between

Threat and Priority Level (r = .576, p < .01). Tracks were selected by the researcher so as to present a variety of threat levels and track types. Significant differences among the tracks was taken as a verification that the selection of the tracks achieved that goal. Priority Levels also changed across Tracks. However, there were not any significant effects or interactions when Priority Level was used as a dependent variable. Threat Levels tended to be higher in the Tense Geopolitical Situation, but the effect was not a significant.

5.0 Content Analysis of Verbal Protocols

Participants were asked to comment on the factors that they considered as they made their threat and priority assessments. Their spoken comments were recorded by audiotape, and later transcribed. The transcripts of the most experienced Naval officers (N = 5; over two years of AD experience) were analyzed by a mixed manifest and latent content analysis, as defined by Berg [1998]. Manifest content are elements that can be directly observed, such as words and phrases. Latent content is an interpretation of the underlying meaning. The verbal protocols of the participants were reduced to 488 statements that were relevant to the research questions. The statements yielded 273 words that corresponded to Factors. There were 22 unique Factors, 21 concepts, and four main themes. Data reduction and definition of factors, concepts, and themes are discussed below. There is a hierarchical relationship among factors, concepts, and themes. Factors are the basic elements, and are organized into concepts, which are organized into themes.

5.1 Identification of Threat Assessment Factors

Any content word that was related to the process of threat assessment was called a *Factor*. It is important to note that an a priori list of factors was not used to search through the protocols. A content word was counted as a Factor if it was explicitly stated, or if it was implied by the word. For example, the comments “high speed” and “350 knots” were counted as instances of one Factor, *Speed*. Factors were recorded and counted for each participant at each time interval and track. In addition, the order in which each factor was mentioned was also recorded.

Participants tended to mention factors in a regular sequence. For example, most participants mentioned that they considered ES Emitter, then Altitude, then Course, and so on. (Note: Order was observed in the data; participants were not asked to list the factors in any particular order.) Table 1 lists the 22 Factors in the standardized order in which they were mentioned. Standardized Order is the Mean Order divided by the number (N) of times a Factor was mentioned. Mean Order is the average order in which the factors were mentioned by the participants. Standardizing reduced the importance of any given factor due simply to the number of times that it was mentioned. Percent is the percentage of times a word was mentioned throughout the data set. Order of mention was also calculated for each Track and each Threat Level.

Table 1. Factors listed in standardized order of mention (Standardized Order).

			Standardized		
Factor	Mean	N	Order	Percent	Description
ES/pltfm	1.98	55	0.04	0.20	ES Emitter or platform type
altitude	2.97	30	0.10	0.11	
course	2.38	21	0.11	0.08	including “Inbound”
speed	3.10	20	0.16	0.07	
airlane	3.11	19	0.16	0.07	Standard commercial route
origin	2.36	14	0.17	0.05	Country of origin
distance	3.47	17	0.20	0.06	Distance from own ship
weapon	3.95	19	0.21	0.07	Weapon type
coord	2.77	13	0.21	0.05	Coordinated activity (formation, targeting)
IFF	3.44	16	0.21	0.06	ID friend or foe mode
wpnenvlp	3.47	15	0.23	0.05	Weapon/threat envelope
CPA	2.43	7	0.35	0.03	Closest point of approach
feetwet	2.86	7	0.41	0.03	Feet wet (over water)
otherid	4.00	5	0.80	0.02	Other ways to ID
launch	3.33	3	1.11	0.01	Missile launch
intel	3.67	3	1.22	0.01	Intelligence reports
support	6.00	4	1.50	0.01	Help from friendly forces
aloft	2.00	1	2.00	0.00	Time in the air
maneuver	2.00	1	2.00	0.00	Number of maneuvers
fuel	4.00	1	4.00	0.00	Fuel remaining
ROE	4.00	1	4.00	0.00	Rules of engagement
visibility	4.00	1	4.00	0.00	weather conditions

5.2 Relationship of Threat Assessment Factors to Threat Level

The factors were analyzed to determine their relationship to Threat and Priority Levels. The analysis yielded three main findings. First, for a given track, participants mentioned only a subset

of the 22 Factors. That is, they used different, but overlapping, sets of Factors to evaluate different Tracks and Threat Levels. Specifically:

- Participants used different Factors for different Tracks (Factor x Track $\chi 2(80) = 126.4$, $p = .001$) and Threat Levels (Factor x Threat Level $\chi 2(48) = 66.5$, $p = .039$).
- Participants mentioned more factors when evaluating threatening Tracks as opposed to nonthreatening Tracks (Factor x Track $\chi 2(105) = 150.3$, $p = .002$).

Second, within a given set of Factors, participants appeared to be evaluating the factors sequentially. ES Emitter or Platform Type were mentioned first by most participants, overall and by track. Table 2 shows the order of consideration, or evaluation, of Factors for threatening and nonthreatening Tracks. The mapping was consistent enough across all participants to produce a significant effect. Specific findings are:

- Participants reviewed the factors in a regular sequence (Spearman r(Factor, Order) = .191, $p = .002$; Factor x Order $\chi 2(128) = 177.7$, $p = .002$).
- Participants mentioned factors in the same order for each Threat Level ($F(3,215)=1.4$, Not Significant).

Finally, the number of factors that were mentioned by participants declined over the course of the scenario ($F(2,68) = 29.4$, $p = .001$). Fewer factors were mentioned at the last data collection interval (15 minutes) than at the first interval. This most likely indicates that participants relied on less information as they became more familiar with a track.

Table 2. Relative sequence of consideration for high and low threat tracks.

Low Threat Track	High Threat Track
ES/platform	ES/platform
altitude	weapon
speed	weapon envelope
airlane	coordinated
coordinated	distance
origin	course
IFF	altitude
course	CPA
	origin
	IFF
	otherID
	feet wet
	speed
	launch
	intel
	support
	airlane

5.3 Concepts and Themes

While reviewing the transcripts for Factors, several consistent clusters of Factors with similar meaning were noted and recorded. The data was scanned several times for these clusters, to ensure consistency. The clusters were called *Concepts*, and are shown in Table 3, organized by Theme. A concept is a unit of meaning, generally a word, phrase, or sentence, that is different from the preceding and following elements in a statement. Clusters are aggregated into *Themes*, which are groups of related concepts. The transcripts were reviewed several times for the themes. Four main themes emerged from the data, and are noted in Table 3. Themes are considered latent content, and are words or phrases that are used to label, and describe, the meaning of several

clusters [Hycner, 1985]. Concepts indicated which factors were used, and their values, while themes highlighted how the factors were related and processed. Concepts and themes were determined from the data, and not generated a priori. Profiles are mentioned in Table 3. They emerged as a way to organize data, and are discussed with the cognitive model.

Table 3. Themes and examples of their subordinate concepts.

Theme 1. Fitting data to a profile – Evaluating data that is consistent with a given profile

- fit-c = commercial air (or non-military)
- fit-m = military

Theme 2. Highlighting unexpected data values

- flag-r = data value is outside of expected range for the primary profile
- flag-t = data value is not typical for the primary profile (may or may not be outside of expected range of values)

Theme 3. Explanations for unexpected data values

- accommodate = explanation of data without assuming intent or changing data values (essentially dismissing the factor)
- alter-v = description of necessary change in data to make it fit the current profile or threat level
- hypothesis = reconciling data to profile by assuming, or hypothesizing, intent

Theme 4. Platform-specific evaluations – Relating data to a specific platform or weapon

- attrib-platform - Naming an attribute or property of a platform
- conclude-platform – A conclusive statement about a platform

Miscellaneous Themes

- reconsider = actions by the track that would cause re-evaluation of the track
- review = restatement of evaluated or known facts
- search = listing of additional, missing evidence that is needed to make an evaluation

---

5.4 Summary of the Verbal Protocol Analysis

It appears that the assumption that Air Defense personnel use a consistent set of factors when assessing aircraft contacts is valid. 22 unique factors were identified, and they were used in a nonrandom manner. That is, the participants were selective regarding the factors that they considered (selected factors depended on threat level), and that the factors were considered in a similar order across all threat levels. In addition, factors were organized into coherent clusters, or themes, which suggests that the participants were systematically using some type scheme to organize and understand the data. The evidence indicates that complex processes are at work. Clearly, a different approach is needed to model the complex process that people bring to the threat assessment process.

6. Cognitive Model of Threat Assessment

Evidence for the model comes largely from the analysis of the hierarchy of factors, concepts, and themes. While the findings that are discussed in the following sections are preliminary, they indicate that participants used aircraft profiles (schemas) to evaluate input data, and assigned Threat Levels that depended on the degree of match between actual input data and expected data (expected by the schema). Therefore, the process of threat assessment appeared to be a simple comparison of factor input data to expected data values. If the data matched expectations, then the resulting Threat Level was low. Otherwise, Threat Level increased as the degree of mismatch increased. The model is consistent with Naturalistic Decision-Making models that use schemas as a way to organize and process information. Profiles, noticing mismatches (inconsistencies), and the role of geopolitical situation are important properties of the cognitive model. Each of these issues is covered, in turn, before the model is described.

6.1 Aircraft Profiles

The majority of each participant’s statements involved fitting data to a limited set of factors. As the analysis of the factors has indicated, sets of factors were fairly stable. They appeared to serve as a means to organize and evaluate data. They formed an outline for expected behaviors of

aircraft. For convenience, and to denote the role they played, these sets of factors were called *Profiles*. The use of schemas, or profiles, as a way to organize factors was not assumed in advance, but rather, it emerged as a result of the content analysis.

A Profile is a schema, in the sense defined by Schank and Abelson [1977]. Like schemas, profiles define the factors to be considered, the sequence of their consideration, and the expected range of data values. These characteristics were clearly evident in the sequencing and grouping of statements in the verbal protocols. For example, the following is a sample from a participant fitting data to his Military profile. Factors such as Feet Wet were not mentioned when participants presumed that a track was a commercial aircraft.

> "Was probable phantom coming out of Iran, either the racetrack type profile, is headed inland, still feet wet."

Participants labeled several types of profiles: Threat, Capable, F-4 (Phantom), Helicopter/Light Air, Commercial Air, and Weapon-Specific (Exocet, Harpoon, Maverick). However, the current data only provide enough samples to justify two profiles: Commercial and Military Aircraft. The Commercial and Military profiles and their expected data values are listed in Tables 4 and 5, respectively. They were created by compiling all of the Fit-m, or Fit-c, statements made by the participants. Note that the Profile concepts of Military and Commercial Air were ascertained from participant's statements, and not from a priori track designations. Fitting a Commercial Air profile is demonstrated by the series of statements made by one participant:

> "Comm air, mode 3, outer air route, same course the whole time."

Table 4. Commercial Air Profile and its expected data values.

Factor	Expected Value
ES emitter	RDR-1500 or none
altitude	moderate to high (20,000-28,000 ft)
speed	moderate to high or constant (350-400 kts)
airlane	on or near (parallel & within 10 nm)
origin	friendly or nonhostile country
IFF	Mode 3 or none
course	steady

Table 5. Military Profile and its expected values.

Factor	Expected Value
origin	hostile
altitude	low-moderate (not too low or too high)
platform	emitter = none or military
weapon	type(s)
weapon envelope	within range of weapon type
platform distance	moderate (not too far or too close)
feet wet	far from land
course	directly inbound
coordinated activity	likely in formation or providing targeting data
speed	moderate-fast
IFF	mode = none or military
airlane	off
DCA or other support	not available
reason to be in area	not plausible
maneuvers	many
visibility	good
CPA	close (within 10 nm)

6.1.1 Expected Data for each Profile

Expected data values for each factor were determined from the participant's statements. For example, the series of statements "Comm air, mode 3, outer air route, same course the whole time." implied that the expected values for a Commercial Aircraft for that participant were IFF mode 3, on airlane, and on a steady course.

An interesting characteristic of the expected values was that each feature appeared to either have a range of input values, or a flexible threshold or trigger value. The exact nature of this characteristic could not be determined from the data. This phenomenon was evidenced by modifying statements such as "very close to the air corridor", "really inbound", and "too low for a commercial airliner". In addition, many values were considered in terms of likelihood or probability (e.g., "possible airlane … that's close enough for an airlane", "believe him to be probable comm air", "possible carry Exocet missile"). Modifiers seemed to highlight the extremeness or unexpectedness of a value. Twenty-five modifiers were recorded. Table 6 shows some of the values, grouped by the factor that they are modifying. Highlighting, or flagging, of data values is a critical process in the model, and is discussed in section 6.2 (Highlighting Atypical Data).

Participants appeared to adjust their sensitivity to the extremeness of incoming data based on the biasing effects of the Geopolitical Situation. Participants who read the tense briefing were more sensitive to deviations from expected behavior (i.e., less deviation from expectations was required to raise threat level) than those who read the relaxed briefing. At a minimum, the effect indicates that factors beyond aircraft flight characteristics must be explicitly considered when modeling threat assessment.

Table 6. Modifiers for expected data values.

Factor	Modifier	Example
Airlane	Close enough	that's close enough for an airlane.
	Near	taken off near the airlane.
	Possible	possible airlane.
	Very close	very close to the air corridor.
	Well off	well off the arrow.
Altitude	Good	altitude 20,000 - that's a good indication.
	Little low	5,000 feet is a little low for that air route.
	Some	he has to lose some altitude before he can shoot.
	Too low	too low for a commercial airliner.
	Very low	he is flying very low at 50 feet.
	Well up	well up in altitude.
Speed	Good	he is moving at a good rate of speed
	Only	but he's only doing 126 knots.
	Real quick	and he's going to be here real quick.
	Relatively slowly	he's coming in relatively slowly.
Weapon Envelope	Almost	actually he's almost in range.
	Already within	he's already within the threat envelope
	Close to inside	getting close to the inside of his range.
	Still within	still within weapons release range.
	Just within	is just within harpoon weapons release range.
	Little	a little farther away
	Well within	well within his weapons release.

6.1.2 Relationship between Aircraft Profile and Threat Level

Threat Levels (1-7) were different for each Profile ($F(2,68) = 6.5$, $p = .003$). Commercial Air profiles were typically rated as least threatening (Mean Threat Level = 1.9), Undetermined

profiles were intermediate (Mean Threat Level = 3.1), and Military profiles were rated as most threatening (Mean Threat Level = 4.0).

It should be noted that this relationship does not always hold. There were cases where participants rated COMM AIR tracks as more threatening than MIL AIR tracks. This result implies that participants were expecting all aircraft (COMM & MIL) to follow typical, but nonaggressive, standard operating procedures (SOPs), and that aircraft became more threatening when they began to deviate from typical SOPs. The nonaggressive nature of the profiles is not an unexpected finding because profiles are built from repeated experience, and most aircraft encountered by AD personnel do not exhibit openly hostile behaviors.

6.2 Highlighting Atypical or Unexpected Data

Sometimes data values were unexpected, unusual, or atypical. Participants would comment on the inconsistency by adding a modifying word or phrase to the data value. Examples are “he’s too high”, “that’s not common”, and “very fast”. This process was called *Flagging* (i.e., participants would flag, or make note of, an unusual value). Occasionally, participants attempted to reconcile unexpected data with their active profile. This process was evidenced when a participant made a flagging statement, and then added a *Hypothesis* or an *Accommodation*. However, most flagged statements were not followed by any type of accommodating or explanation statement. The following statement is an example of a participant flagging a Military IFF value of Mode 2, and then Accommodating it into his active Commercial Air profile:

> “He has mode 2, which is unique for this area, but it’s a possibility that I could be picking it up from somewhere else.”

At times, it appeared that participants discounted flagged factors, either through accommodation or giving evidence to the contrary, such as“… but it's not uncommon to see commercial air flying across at that [altitude].”, from a participant who had just commented on the low altitude of a track. Occasionally, participants would elaborate, or provide supporting information, about a factor, as in “… Helos are difficult to cover with radar.”

Elaborations seemed to increase the reliance on, or weight of, a factor. However, the evidence did not support the notion that factors were weighted, either up or down. The number of Accommodate and Hypothesize statements were not related to Threat Level, although they demonstrated similar trends to Flagging

6.2.1 Relationship between Flagging and Threat Level

The analysis of flagging and explaining statements indicated that inconsistency played a major role in determining Threat Level, and that its affect was unidirectional: More inconsistency lead to higher Threat Levels. Inconsistentcy was indicated by a flagging statement. As the Number of Flagged statements increased, so did Threat Level ($F(3,67) = 4.5$, $p = .006$). The Number and Proportion of Flagged statements were significantly correlated with Threat Level ($r = .355$, $p < .05$). The Proportion of Flagged statements was also related to the type of Profile that the participants were using. Most flagged statements occurred with the Unknown Profile, then Military, and then Commercial Air. However, the relationship between Flagging and Profile was not significant.

6.3 The Role of Geopolitical Situation Bias

While earlier analysis indicated that the Geopolitical Situation Bias had no affect on Threat Level, it clearly biased the process of threat assessment. Participants flagged factors as extreme or inconsistent more often after reading the Tense briefing than after reading the Relaxed briefing ($F(1,39) = 5.7$, $p = .019$). The Number and Proportion of Flagged statements were significantly correlated with Geopolitical Situation Bias ($r = -.277$, $p < .05$) across all tracks and data collection intervals. Participants tended to make more Hypothesis statements in the Tense condition ($F(1,9) = 5.0$, $p = .052$), while the number of Accommodating statements were similar for both groups. These findings are summarized in Figure 3.

The modifiers that participants used in flagging statements suggested that the Geopolitical Situation Bias affected the range of acceptable input values. In the Tense condition, for example,

military aircraft that were anywhere close to their weapon release envelope were flagged (e.g., "… he's almost in range."). While in the Relaxed condition, most military aircraft had to at least be within (usually well within) their weapon release envelope before they were flagged (e.g., "… he's within his envelope now.").

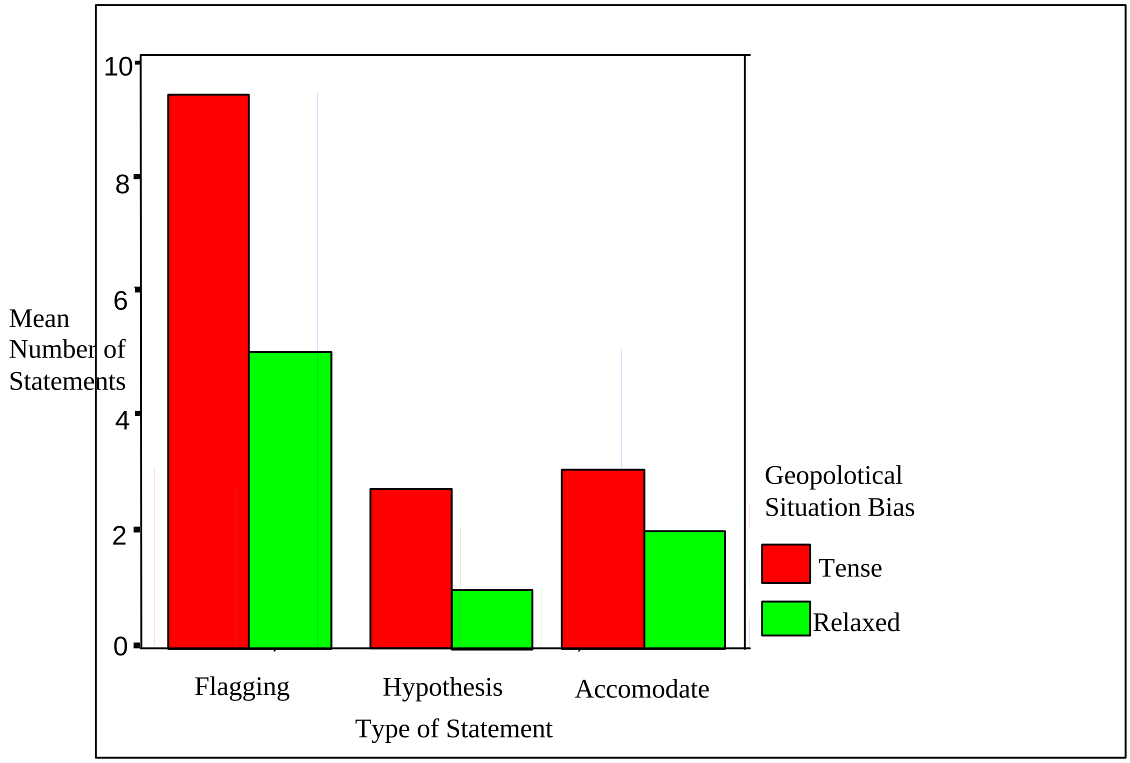

Figure 3. Average number of statements in each Geopolitical Situation Bias condition.

It appeared that Geopolitical Situation was altering a threshold or trigger value, but the data are only suggestive at this point. Even though the details were elusive, the results suggest that Geopolitical Situation played an indirect role in determining Threat Level in the following manner:

- Geopolitical Situation reduced a participant's tolerance for deviations from expected behavior. Tense Situation = Less Tolerance (Most deviations from expectations were Flagged). Relaxed Situation = More Tolerance (Only major deviations from expectations were Flagged).
- Less tolerance for deviation lead to more mismatches between expected and the actual data values.
- More mismatches resulted in higher Threat Levels

7. Specifications for a Cognitively Based Model of Threat Assessment

The data indicate that expert Air Defense officers use profiles to evaluate data, that the geopolitical situation affects their expectations, and that threat level of an aircraft depends on the degree of match between expectations and actual input data values. Therefore, the key feature of the proposed model is that Threat Level is built from an evaluation process that compares the degree of fit between input data and profile generated expectations. The model is shown in Figure 4. It represents a preliminary step in the creation of an operational threat assessment system. It accurately reflects the processes that were evident in the verbal data, and the assignment of Threat Level. The cognitive model has also identified key features which must be evaluated for future situation models, and has provided a framework for future testing, evaluation, and refinement.

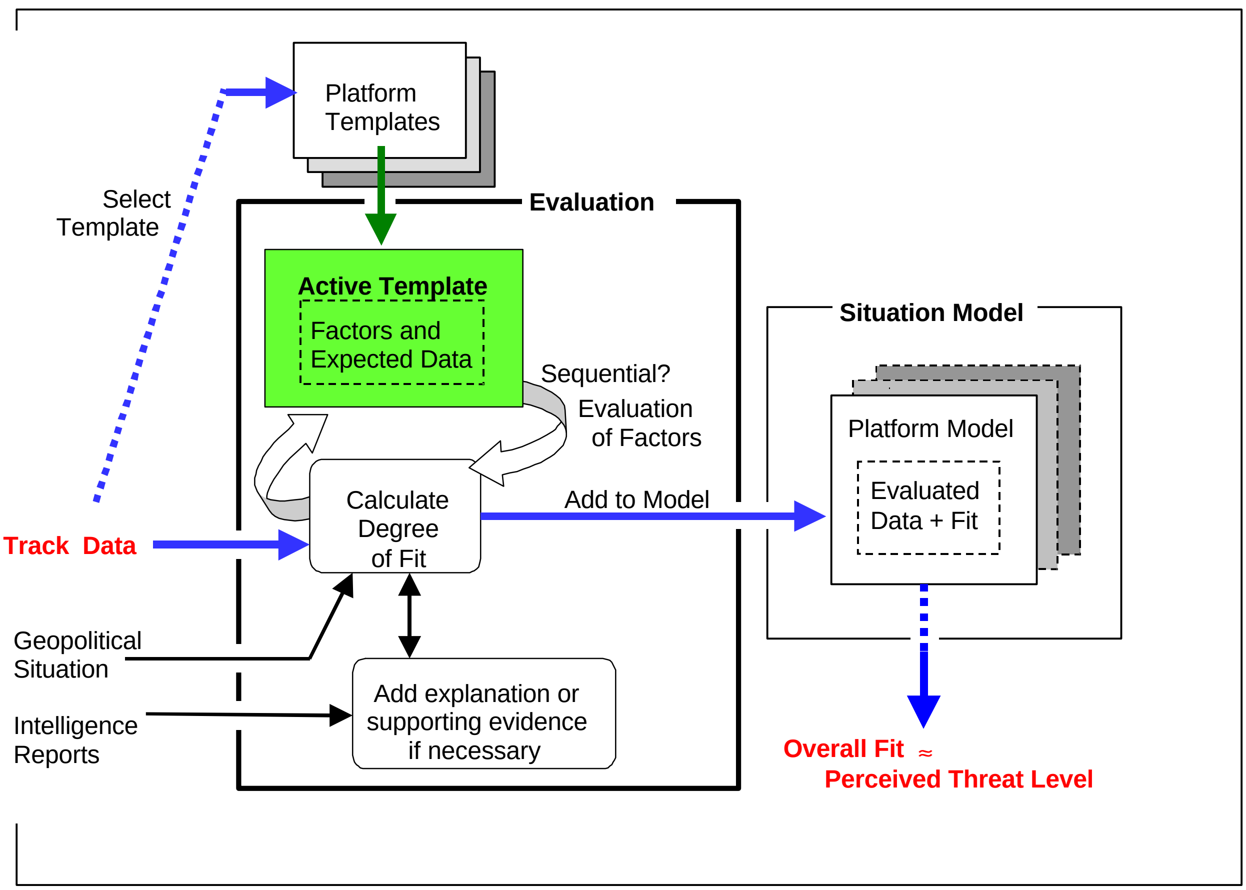

Figure 4. Cognitively-Based Model of Threat Assessment.

7.1 Information Flow through the Model

The steps in the *Evaluation* phases were accumulated and abstracted across all participants, tracks, and times, and were represented to some degree in every assessment interval by every participant. For a given track, a profile is selected and activated. Participants seemed to activate a profile that corresponded to the emitter data. The verbal protocol data do not indicated how this was done. The active profile determined which factors were evaluated, and participants overwhelmingly chose to accommodate or explain unusual data, rather than to switch profiles (only 3.7% of all participant's profile fitting statements indicated that they were considering an alternative profile). Participants appeared to compare factors, in order, against profile expectations. They flagged some inconsistent values, and occasionally offered explanations in an attempt to make the data fit their expectations. Some participants added supporting or counter evidence based on personal experience. The situation model was a mechanism for accumulating the degree of fit across all of the evaluated factors. It was added because the data indicated that Threat Level was a function of the overall degree of fit to expectations. The accumulated degree of fit is the Threat Level for the given track.

7.2 Calculating the Degree of Fit between Data and Expectations

A key assumption of the model was that the degree of fit, or correspondence, between actual and expected data values for a given track, was indicated, albeit roughly, by the Proportion of Flagged statements. The relationship between Proportion of Flagged statements and Threat Level is shown in Figure 5: Threat Level increased as the Proportion of Flagged statements increased. Proportion of Flagged statements is an imprecise indicator of Degree of Fit because one cannot assume that participants mentioned or flagged all of the factors that they were considering.

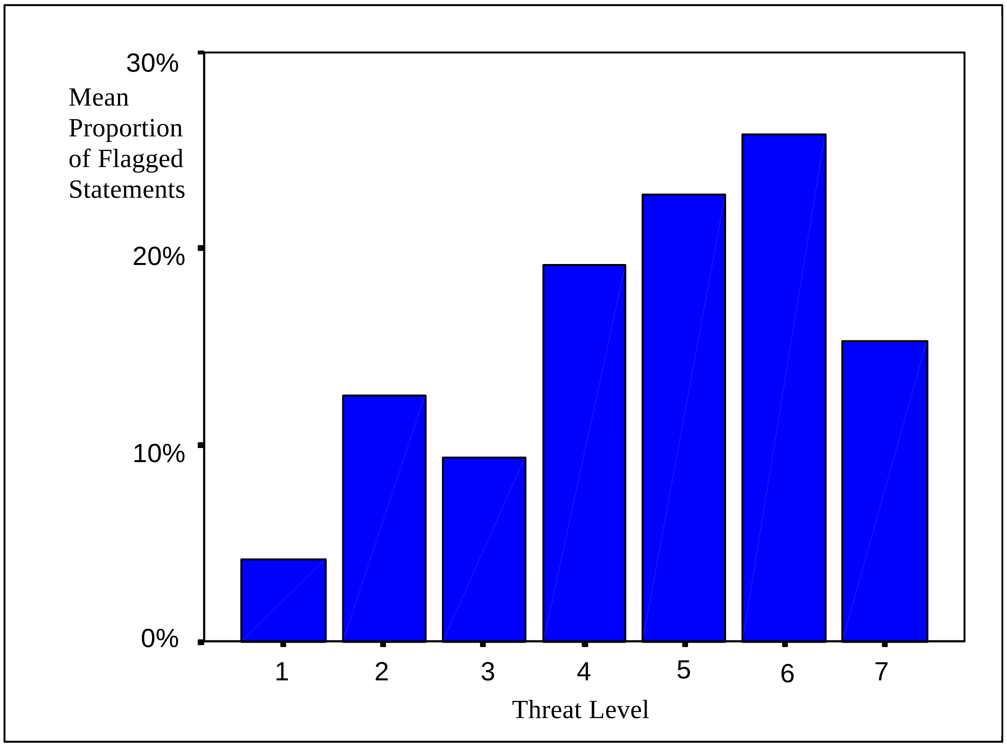

Figure 5. The relationship between Proportion of Flagged statements and Threat Level.

Degree of Fit was selected as a measure of correspondence for two reasons. First, participants statements indicated that they were considering flexible ranges of expected data (e.g., “too high”, “really feet wet”, “very fast”). Ranges imply degrees rather than absolutes, and that they were “fitting” the data to their expectations. Second, the degree of correspondence between a mental model and reality is generally regarded as a mechanism for evaluating the validity of mental models [Johnson-Laird, 1983]. The better the data can be mapped onto the model, the better the fit, and the more valid the model. In addition, there is evidence for “fitting” in relation to situation awareness. Adams, Tenney, and Pew [1995] noted that a mismatch between the perceived and actual situation was a major factor in many aircraft accidents.

7.3 Calculating Threat Level

In the model, each factor is evaluated, in order, against an expected value. The results of the content analysis suggests that Threat Level was not a one-time calculation, but was constantly being modified as factors were being evaluated. However, it does not appear that threat level is

simply the accumulation, or summing, of evidence. If this were the case, then one would expect that the number of factors would be related to Threat Level (i.e., more threatening factors leads to higher Threat Levels); it was not.

At this time, the data tend to favor the conclusion that Threat Level is inversely related to the Degree of Fit between an input value and a profile generated expected value, and is mediated by Geopolitical Situation Bias. For example, a commercial airliner is expected to have an altitude of over 20,000 ft. In a relaxed Geopolitical Situation, a track with an altitude of 28,000 ft is flying within expectations, and may be assigned Threat Level 1. An altitude of 18,000 ft could raise the Threat Level of the track to 3, and an altitude of 9,000 ft would raise the Threat Level even more, perhaps to 5. In a tense Geopolitical Situation, the Threat Levels would increase, perhaps to 2, 4, and 6, for the same altitudes as in the relaxed situation.

8. Discussion

The goals of this project were to describe the data and processes that were used by experienced air defense tactical decision-makers as they assessed and prioritized aircraft tracks, and to specify a cognitively-based threat assessment model of that process. Data were collected from experienced U.S. Navy air defense officers as they interacted with a realistic scenario. Participants assigned threat and priority levels to selected aircraft in the vicinity of their ship, and described the data and rationale that they used to arrive at their decisions.

The findings showed that when experienced Air Defense officers evaluate the Threat Level of an aircraft, they:

- Quickly activate an aircraft Profile.
- Sequentially evaluate input data against expectations.
- Are more likely to notice deviations from expected data values in Tense Geopolitical Situations.
- Explain and accommodate deviations rather than switch schemas.
- Set a Threat Level that is directly related to the fit between data and expectations.
    - Poor fit between data and expectations = Higher Threat Level.

- Good fit between data and expectations = Lower Threat Level.

Profiles are schemas that specify the expected behaviors of a class of aircraft. They define the factors (e.g., speed) to be considered, the sequence of their consideration, and the expected data values (e.g., Commercial IFF = 3). Participants would sometimes mention that a data value was unexpected or atypical (e.g., low altitude). This was taken as evidence that they were fitting data to expectations. Threat Level was inversely related to the degree of fit of data to expectations. The better the fit to expectations, the lower the Threat Level. Participants sensitivity to deviations from expectations depended on the geopolitical situation. They were more likely to notice small discrepancies in a tense situation than in a relaxed situation. The processing flow was described in the Threat Assessment model.

Results from the present study suggest that threat assessment is similar to situation assessment in the NDM framework in that Profiles are used evaluate information and to make judgments, and explanations are given when data are atypical. In addition, the model appears to be closely related to the early stages of processing of other models . However, there are some interesting differences. The model opens up the black box of “situation recognition”, and offers some insight into how situation awareness is built. For example, it does not appear that assessment relies on matching a pattern of data to a series of profiles until a match is found; as in NDM models. Instead, participants seemed to activate only one profile, and then use it to evaluate input data, regardless of the profile’s appropriateness. Adams, Tenney, and Pew [1995] came to a similar conclusion regarding situation awareness in pilots; who sometimes overlooked or minimized evidence that was not consistent with their beliefs about the situation. Current findings also indicated that the evaluation process is sequential. However, most current NDM models assume that processing is parallel.

Despite the preliminary nature of this study, many key issues were identified that will provide guidance for future model development. However, because participant’s responses were limited to the characteristics of the aircraft within the experimental scenario, it also raised several questions, such as: Are the profiles robust?, What are the specific values associated with each attribute?, Is the sequential nature of attribute evaluation an artifact of the data collection

process?, and what degree of mismatch between data and expectations is required to get participants to switch profiles? A series of studies are in progress that will attempt to verify and expand upon these results. The first study will refine the set of attributes and their corresponding data values. Following studies are designed to further evaluate the effect of bias on the assessment process, and to determine how participants process attributes (parallel, weighted, or sequentially).

9. References

[Berg, 1998] Berg, B.L. *Qualitative methods for the social sciences*. New York: Allyn and Bacon. 1998.

[Cohen et al., 1993] Cohen, M.S., Adelman, L., Tolcott, M.A., Bresnick, T.A., & Marvin, F. A cognitive framework for battlefield commander's situation assessment. (Tech. Rpt. 93-1). Cognitive Technologies, Inc., 4200 Lorcom Lane, Arlington, VA 22207. 1993.

[Cohen, Freeman, & Thompson, 1997] Cohen, M.S., Freeman, J.T., & Thompson, B.B. Training the naturalistic decision maker. In CE Zsambok & G Klein (Eds.), *Naturalistic Decision Making* (pp. 257-268). Mahwah, NJ: Lawrence Erlbaum Associates, Inc. 1997.

[Cohen & Klein, 1995] Cohen, M.S. & Klein, G.A. Training cognitive skills for tactical planning and situation assessment. Presented at a Meeting of the Technical Advisory Board Program on Tactical Decision Making Under Stress, San Diego, CA, 28 November. Cognitive Technologies, Inc., 4200 Lorcom Lane, Arlington, VA 22207. 1995.

[Endsley, 1995] Endsley, M.R. Toward a theory of situation awareness in dynamic systems. *Human Factors, 37,* 32-64. 1995.

[Federico, 1997] Federico, P-A. An empirical examination of metacognitive models of situation assessment. *Human Factors*, *39*, 149-157. 1997.

[Hycner, 1985] Hycner, R.H. Some guidelines for the phenomenological analysis of interview data. *Human Studies, 8*, 279-303. 1985.

[Johnson-Laird, 1983] Johnson-Laird, P. *Mental Models*. Cambridge, MA: Harvard University Press. 1983.

[Kelly & Moore, 1996, December]. Kelly, R.T. & Moore, R.A. *Situational parameters and scenario events associated with decision making workload*. (TADMUS Tech. Note). Pacific Science & Engineering Group, Inc., 6310 Greenwich Drive, Suite 200, San Diego, CA 92122. 1996.

[Klein, 1993]. Klein, G.A. A recognition-primed decision (RPD) model of rapid decision making. In GA Klein, J Orasanu, R Calderwood, & C Zsanbok (Eds.), *Decision making in action: Models and methods* (pp. 138-147). Norwood, NJ: Ablex Publishing Corp. 1993.

[Klein, 1997]. Klein, G.A. The recognition-primed decision (RPD) model: Looking back, looking forward. In CE Zsambok & G Klein (Eds.), *Naturalistic Decision Making* (pp. 285-292). Mahwah, NJ: Lawrence Erlbaum Associates, Inc. 1997.

[Lipshitz & Shaul, 1997]. Lipshitz, R. & Shaul, O.B. Schemata and mental models in recognition-primed decision making. In CE Zsambok & G Klein (Eds.), *Naturalistic Decision Making* (pp. 293-303). Mahwah, NJ: Lawrence Erlbaum Associates, Inc. 1997.

[Marshall et al., 1996]. Marshall, S.P., Christensen, S.E., & McAllister, J.A. Cognitive differences in tactical decision making. In *Proceedings of the 1996 Command and Control Research and Technology Symposium* (pp. 122-132). Monterey, CA: Naval Postgraduate School. 1996.

[Schank & Abelson, 1977]. Schank, R. & Abelson, R. *Scripts, plans, goals, and understanding*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc. 1997.

[Serfaty et al., 1997]. Serfaty, D., MacMillan, J., Entin, E.E., & Entin, E.B. The decision-making expertise of battlefield commanders. In CE Zsambok & G Klein (Eds.), *Naturalistic Decision Making* (pp. 233-246). Mahwah, NJ: Lawrence Erlbaum Associates, Inc. 1997.

[Smith & Marshall, 1997]. Smith, D.E. & Marshall S.P. Applying hybrid models of cognition to decision aids. In CE Zsambok & G Klein (Eds.), *Naturalistic Decision Making* (pp. 331-342). Mahwah, NJ: Lawrence Erlbaum Associates, Inc. 1997.

[Sonalysts, Inc., 1995]. Sonalysts, Inc. *TADMUS experiment: Decision-making facility for tactical teams (DEFTT) scenario documentation*. Sonalysts, Inc., 5675 Ruffin Road, Suite 210, San Diego, CA 92123. 1995.

[Zachary et al., 1992]. Zachary, W.W., Zaklad, A.L., Hicinbothom, J.H., Ryder, J.M., Purcell, J.A., & Wherry, Jr., J.R. *COGNET representation of tactical decision-making in ship-based anti-air warfare.* (Tech. Rpt. 920211.9009). CHI Systems, Inc., Gwynedd Plaza III, Bethlehem Pike at Sheble Lane, Spring House, PA 19477. 1992.

[Zsambok & Klein, 1997]. Zsambok, C.E. & Klein, G.A., (Eds.). *Naturalistic Decision Making*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc. 1997.